yang_seonhyeHW15

Seonhye Yang

2/23/2019

```r
install.packages("alr4")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/3.5'
## (as 'lib' is unspecified)
```

```r
install.packages("lmtest")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/3.5'
## (as 'lib' is unspecified)
```

```r
library(alr4)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: effects
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```r
library(MASS)
library(data.table)
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

Question 1

```r
attach(jevons, warn.conflicts = F)
fit <- lm(Weight~Age)
summary(fit)
```

```
##
## Call:
## lm(formula = Weight ~ Age)
##
## Residuals:
##        1        2        3        4        5
## -0.00204  0.00107  0.00368 -0.00241 -0.00030
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.9998500  0.0030123 2655.74 1.18e-10 ***
```

```
## Age         -0.0253100  0.0009082  -27.87 0.000101 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002872 on 3 degrees of freedom
## Multiple R-squared:  0.9962, Adjusted R-squared:  0.9949
## F-statistic: 776.6 on 1 and 3 DF,  p-value: 0.0001014
```

```
confint(fit, level = .95)
```

```
##                   2.5 %      97.5 %
## (Intercept)  7.99026354  8.00943646
## Age         -0.02820043 -0.02241957
```

##Question 2

```
fit1 <- lm(Weight~Age, weights=n)
fit1
```

```
##
## Call:
## lm(formula = Weight ~ Age, weights = n)
##
## Coefficients:
## (Intercept)          Age
##     7.99822     -0.02469
```

```
summary(fit1)
```

```
##
## Call:
## lm(formula = Weight ~ Age, weights = n)
##
## Weighted Residuals:
##         1         2         3         4         5
## -0.011442  0.012903  0.019537 -0.013415 -0.008627
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.9982244  0.0020332  3933.9 3.62e-11 ***
## Age         -0.0246927  0.0008427   -29.3 8.73e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01764 on 3 degrees of freedom
## Multiple R-squared:  0.9965, Adjusted R-squared:  0.9954
## F-statistic: 858.6 on 1 and 3 DF,  p-value: 8.729e-05
```

```
confint(fit, level = .95)
```

```
##                   2.5 %      97.5 %
## (Intercept)  7.99026354  8.00943646
## Age         -0.02820043 -0.02241957
```

```
w <- diag(fit1$weights)
w
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  123    0    0    0    0
```

```
## [2,]    0   78    0    0    0
## [3,]    0    0   32    0    0
## [4,]    0    0    0   17    0
## [5,]    0    0    0    0   24
```

```r
sigma <- solve(w)
sigma
```

```
##             [,1]       [,2]    [,3]       [,4]       [,5]
## [1,] 0.008130081 0.00000000 0.00000 0.00000000 0.00000000
## [2,] 0.000000000 0.01282051 0.00000 0.00000000 0.00000000
## [3,] 0.000000000 0.00000000 0.03125 0.00000000 0.00000000
## [4,] 0.000000000 0.00000000 0.00000 0.05882353 0.00000000
## [5,] 0.000000000 0.00000000 0.00000 0.00000000 0.04166667
```

Question 3

```r
fit2 <- lm(Weight~Age, weights=1/SD^2)
confint(fit2, level = .95)
```

```
##                   2.5 %      97.5 %
## (Intercept)  7.99171035  8.00366617
## Age         -0.02736101 -0.02168001
```

```r
summary(fit2)
```

```
##
## Call:
## lm(formula = Weight ~ Age, weights = 1/SD^2)
##
## Weighted Residuals:
##        1        2        3        4        5
## -0.04739  0.07274  0.10138 -0.08396 -0.03896
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.9976883  0.0018784 4257.71 2.86e-11 ***
## Age         -0.0245205  0.0008926  -27.47 0.000106 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09378 on 3 degrees of freedom
## Multiple R-squared:  0.996,  Adjusted R-squared:  0.9947
## F-statistic: 754.7 on 1 and 3 DF,  p-value: 0.0001059
```

```r
w2 <- diag(fit2$weights)
w2
```

```
##         [,1]    [,2]     [,3]     [,4]    [,5]
## [1,] 5037.07    0.00   0.0000   0.0000   0.000
## [2,]    0.00 1937.24   0.0000   0.0000   0.000
## [3,]    0.00    0.00 851.9719   0.0000   0.000
## [4,]    0.00    0.00   0.0000 607.5611   0.000
## [5,]    0.00    0.00   0.0000   0.0000 348.984
```

```r
sigma2 <- solve(w)
sigma2
```

```
##             [,1]       [,2]    [,3]       [,4]       [,5]
## [1,] 0.008130081 0.00000000 0.00000 0.00000000 0.00000000
## [2,] 0.000000000 0.01282051 0.00000 0.00000000 0.00000000
## [3,] 0.000000000 0.00000000 0.03125 0.00000000 0.00000000
## [4,] 0.000000000 0.00000000 0.00000 0.05882353 0.00000000
## [5,] 0.000000000 0.00000000 0.00000 0.00000000 0.04166667
```

```r
lpga2009_1 <- read.csv("/cloud/project/lpga2009.csv")
lpga2009 <- data.table(lpga2009_1, Header = T)
```

Question 4

```r
attach(lpga2009, warn.conflicts = F)
lpga <- lm(prize~percentile)
lpga_coef = lpga$coefficients
lpga_coef
```

```
## (Intercept)  percentile
##  -785350.20    20366.62
```

```r
bptest(lpga)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  lpga
## BP = 24.895, df = 1, p-value = 6.055e-07
```

$\chi^2_{BP} > \chi^2_{\alpha}(p)$ so therefore, we can reject the null hypothesis.

```r
ln_prize<-log(prize)
ln_percentile<- log(percentile)
lpga_ln <- lm(ln_prize~ln_percentile)
lpga_coef_ln = lpga_ln$coefficients
lpga_coef_ln
```

```
##   (Intercept) ln_percentile
##     -4.951524      4.262748
```

```r
bptest(lpga_ln)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  lpga_ln
## BP = 5.5482, df = 1, p-value = 0.0185
```

$\chi^2_{BP} > \chi^2_{\alpha}(p)$ so therefore, we can reject the null hypothesis.

Question 5

The linear model does not satisfy Gauss-Markov as Gauss-Markov states that $V(\epsilon_i) = \sigma^2$, which is not true in the given case.

We can transform this:

$$y = \mathbf{X}\beta + \epsilon$$
$$y(\Sigma^{-1})^{\frac{1}{2}} = \mathbf{X}\beta(\Sigma^{-1})^{\frac{1}{2}} + \epsilon(\Sigma^{-1})^{\frac{1}{2}}$$
$$y' = \mathbf{X}\beta' + \epsilon'$$

Now our $V(\epsilon') = \sigma^2\Sigma$